

# NONPARAMETRIC ESTIMATION OF DISTRIBUTION FUNCTIONS OF NONSTANDARD MIXTURES

Alan M. Polansky, Division of Statistics, Northern Illinois University, De Kalb, IL 60115.

## SUMMARY

Nonstandard mixtures are those that result from a mixture of a discrete and a continuous random variable. They arise in practice, for example, in medical studies of exposure. Here, a random variable that models exposure might have a discrete mass point at no exposure, but otherwise may be continuous. In this paper we explore estimating the distribution function associated with such a random variable from a nonparametric viewpoint. We assume that the locations of the discrete mass points are known so that we will be able to apply a classical nonparametric smoothing approach to the problem. The proposed estimator is a mixture of an empirical distribution function and a kernel estimate of a distribution function. A simple theoretical argument reveals that existing bandwidth selection algorithms can be applied to the smooth component of this estimator as well. The proposed approach is applied to two example sets of data.

*Keywords:* Auditing; Bandwidth estimation; Empirical distribution function; Exposure data; Kernel; Plug-in estimate

## 1. INTRODUCTION

In practice one often encounters data that behaves in a continuous manner except at a single mass point. Examples where such random variables arise in practice come from many areas of statistical analysis including accounting, biostatistics, environmental statistics and reliability. For example, one is often interested in modeling the amount of exposure to a certain toxin a subject has experienced. In this case it is often possible that there is a non-zero probability that the subject has not been exposed to the toxin at all. Hence, there is a discrete mass point at 0. Otherwise, the random variable may behave in a continuous manner. In biostatistics such a distribution arises when one considers exposure to environmental tobacco smoke where the variable of interest is the average number of cigarettes smoked per day in a household. This data will have a mass point at zero, but may behave in a continuous manner otherwise. For an example of this type of data see Rubin

(1990). In accounting, the distribution of errors detected by an audit will have a mass point at zero indicating entries that are correct, but will be virtually continuous otherwise. Further examples of this type of data are discussed by Panel on Nonstandard Mixtures of Distributions (1989).

Because neither a density or a mass function correctly captures the true structure of such data, the distribution function of such a random variable is often of interest. In practice an estimate of the distribution function can be used to estimate the probability that a random variable is greater than a specified value. For example, in environmental statistics or biostatistics, such an estimate could correspond to the probability that exposure to a toxin exceeds an established safety level. Further, an estimate of a distribution function can be used to estimate quantiles of a nonstandard mixture.

The first approach to dealing with a continuous distribution that has a mass point at 0 was apparently presented by Aitchison (1955) who considers efficient parameter estimation. Distributions of standardized statistics calculated on data from nonstandard mixtures is considered by Vännman (1995). Schucany and Wang (1994) discuss methods for bootstrapping nonstandard mixtures. An excellent introduction to problems of this type that arise in auditing is given by Panel on Nonstandard Mixtures of Distributions (1989). This paper also provides an extensive annotated bibliography.

In this paper we will address the issue of nonparametrically estimating the distribution function of a nonstandard mixture based on a random sample. We will assume that the set of jump points is known. In Section 2 we explore the properties of two estimators of such a distribution function. The first is the usual empirical distribution function and the second is a mixture of an empirical distribution function and a kernel estimate of a distribution function. We address the theoretical properties of these two estimators as well as the question of bandwidth selection for the smooth component of the mixture estimate. In Section 3 we apply the two estimators to two example sets of data. A general discussion is presented in Section 4. Details of the technical arguments for the theoretical results are given in the Appendix.

## 2. ESTIMATORS

To develop our estimators in a statistical framework consider a set of independent and identically distributed random variables, denoted  $X_1, \dots, X_n$ , following a distribution  $F$ . According to Theorem 1.2.3 of Chung (1974), the distribution function  $F$  can be uniquely decomposed into two functions as  $F(x) = F_d(x) + F_c(x)$ , for all  $x \in \mathbb{R}$  where  $F_c$  is positive, increasing, and continuous everywhere,

and  $F_d$  is a nondecreasing step function with a countable set of jump points denoted by  $A$ . We will assume that both of these functions are bounded between 0 and 1, and for the moment that  $F_d \not\equiv 0$  and  $F_c \not\equiv 0$ . Now define

$$\alpha = \lim_{x \rightarrow \infty} F_d(x),$$

so that  $\alpha \in (0, 1)$ . An equivalent representation of  $F$  is given as the unique convex combination

$$F(x) = \alpha F_1(x) + (1 - \alpha) F_2(x),$$

where  $F_1(x) = F_d(x)/\alpha$  and  $F_2(x) = F_c(x)/(1 - \alpha)$  for all  $x \in \mathbb{R}$ . In the case where  $F_d \equiv 0$ , then  $F \equiv F_c$  and we take  $\alpha = 0$ . Similarly, when  $F_c \equiv 0$ , then  $F \equiv F_d$  and we take  $\alpha = 1$ . Thus when  $\alpha = 0$ ,  $F$  is a distribution function for a continuous random variable and when  $\alpha = 1$ ,  $F$  is a distribution function for a discrete random variable. When  $\alpha \in (0, 1)$ , then  $F$  is a distribution function for a random variable that is discrete with probability  $\alpha$  and is continuous with probability  $1 - \alpha$ .

Our goal is to use  $X_1, \dots, X_n$  to nonparametrically estimate  $F$ . A common estimator is the empirical distribution function. Let  $\Omega(x; A)$  be the indicator function defined by

$$\Omega(x; B) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B. \end{cases}$$

The empirical distribution function, denoted  $\hat{F}$ , is defined by

$$\hat{F}(x) = n^{-1} \sum_{i=1}^n \Omega(x; [X_i, \infty)),$$

which is a step function with a step of height  $n^{-1}$  at each observed sample point  $X_i$ . When considered as a pointwise estimator of  $F$ ,  $\hat{F}(x)$  is an unbiased and strongly consistent estimator of  $F(x)$ . In a global sense, if  $D_n = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F(x)|$ , the Glivenko-Cantelli Theorem implies that  $D_n \xrightarrow{a.e.} 0$  as  $n \rightarrow \infty$ . See Section 2.1 of Serfling (1980) for details of these results. Note, however, that  $\hat{F}$  is a distribution function for a discrete random variable, and hence will not reflect the true nature of the structure of the nonstandard mixture.

To motivate an estimate that does reflect the true structure of the nonstandard mixture, consider estimating  $\alpha$ ,  $F_1$ , and  $F_2$  separately. An obvious estimate of  $\alpha$  is

$$\hat{\alpha} = n^{-1} \sum_{i=1}^n \Omega(X_i; A), \tag{1}$$

which is the proportion of observations in the sample that are in  $A$ . To estimate  $F_1$  we use the empirical distribution function. Noting that we wish to compute this estimate only on the sample values in  $A$ , we can write this estimate as

$$\hat{F}_1(x) = \left\{ \sum_{i=1}^n \Omega(X_i; A) \right\}^{-1} \left\{ \sum_{i=1}^n \Omega(x; [X_i, \infty)) \Omega(X_i; A) \right\}. \quad (2)$$

The estimate of  $F_2$  should be continuous. A nonparametric method for constructing such an estimate of  $F_2$  uses a kernel estimate of a distribution function. Let  $k$  be a continuous function such that  $\mu_0(k) = 1$ ,  $\mu_1(k) = 0$  and  $0 < \mu_2(k) < \infty$ , where

$$\mu_l(k) = \int_{-\infty}^{\infty} t^l k(t) dt.$$

With these assumptions  $k$  can be taken to be a continuous density that is symmetric about zero with variance  $\mu_2(k)$ . Define

$$K(x) = \int_{-\infty}^x k(t) dt,$$

the distribution function of the density  $k$ . Now consider a random sample  $Y_1, \dots, Y_n$  from a continuous distribution  $G$ . The well known kernel estimate of  $G$  is

$$\hat{G}(y; h) = n^{-1} \sum_{i=1}^n K\{(y - Y_i)/h\}.$$

This estimator was apparently first proposed by Nadaraya (1964). The parameter  $h$  is a smoothing parameter called a bandwidth. The function  $K$  is called a kernel function. Typically  $h$  is chosen to minimize some measure of performance. We will consider this choice below. The choice of  $K$ , at least from an asymptotic viewpoint, is generally not as important as the choice of  $h$ , though an optimal form of the kernel does exist. See Jones (1990). For the remainder of the paper we will assume that  $K$  is a standard normal distribution function.

We now apply this estimator to our situation. The kernel estimate of  $F_2$  is given by

$$\hat{F}_2(x; h) = \left\{ \sum_{i=1}^n \Omega(X_i; \mathbb{R} \setminus A) \right\}^{-1} \left[ \sum_{i=1}^n K\{(x - X_i)/h\} \Omega(X_i; \mathbb{R} \setminus A) \right], \quad (3)$$

where the notation  $B \setminus C$  denotes the set of elements in  $B$ , that are not in  $C$ . Combining the estimators given in Equations (1) - (3) yields

$$\hat{F}(x; h) = \hat{\alpha} \hat{F}_1(x) + (1 - \hat{\alpha}) \hat{F}_2(x; h) = n^{-1} \sum_{i=1}^n \tilde{\Omega}_i(x; A) \quad (4)$$

where

$$\tilde{\Omega}_i(x; A) = \begin{cases} \Omega(x; [X_i, \infty)) & \text{when } X_i \in A, \\ K\{(x - X_i)/h\} & \text{when } X_i \in \mathbb{R} \setminus A. \end{cases}$$

To compare the empirical distribution function to the smoothed estimate in a theoretical setting we will use the usual measure of performance from the smoothing literature called the mean integrated square error, given by

$$\text{MISE}(\hat{F}, F) = \int_{-\infty}^{\infty} E\{\hat{F}(x) - F(x)\}^2 dx = \int_{-\infty}^{\infty} \text{MSE}(x; \hat{F}, F) dx, \quad (5)$$

where  $\text{MSE}(x; \hat{F}, F)$  is the pointwise mean squared error of the estimator  $\hat{F}(x)$ . For the empirical distribution function it can be shown that

$$\text{MISE}(\hat{F}, F) = n^{-1} \int_{-\infty}^{\infty} F(x)\{1 - F(x)\} dx. \quad (6)$$

To obtain the mean integrated square error for the estimator given in Equation (4) we make the following assumptions on  $F_2$ . Assume that  $F_2$  is differentiable everywhere and that  $F_2'$  is continuous and differentiable with a finite mean and has a square integrable derivative. Under these conditions the mean integrated square error of  $\hat{F}(x; h)$  is given in Theorem 1 below.

*Theorem 1.* Under the assumptions of this section, when  $\hat{F}$  is the smoothed estimator given in Equation (4),

$$\text{MISE}(\hat{F}, F) = n^{-1} \int_{-\infty}^{\infty} F(x)\{1 - F(x)\} dx - n^{-1} h(1 - \alpha)C_1 + h^4(1 - \alpha)^2 C_2/4 + o(h^4 + n^{-1}h), \quad (7)$$

where

$$C_1 = \int_{-\infty}^{\infty} K(x)\{1 - K(x)\} dx,$$

and

$$C_2 = \mu_2^2(k) \int_{-\infty}^{\infty} \{F_2''(x)\}^2 dx.$$

To derive an optimal expression for the bandwidth we asymptotically minimize Equation (7) with respect to  $h$ . This minimization yields

$$h_0 = \left\{ \frac{C_1}{n(1 - \alpha)C_2} \right\}^{1/3}, \quad (8)$$

as the optimal value of the bandwidth. The value of  $\text{MISE}(\hat{F}_{n,h})$  when using  $h = h_0$  is

$$\text{MISE}_0(\hat{F}_{n,h}) = n^{-1} \int_{-\infty}^{\infty} F(x)\{1 - F(x)\} dx - 3n^{-4/3}(1 - \alpha)^{2/3} C_1^{4/3} C_2^{-1/3}/4 + o(n^{-4/3}). \quad (9)$$

One can observe from Equation (9) that  $\text{MISE}(\widehat{F}, F) = O(n^{-1})$ . This is the same rate achieved by the empirical distribution function so that smoothing the continuous part of  $F$  only has a second-order effect on the mean integrated square error. Hence our smoothed estimate loses little efficiency, at least from an asymptotic viewpoint, when compared to the empirical distribution function. Indeed, in the finite sample case, we may slightly improve performance under some circumstances. Unfortunately  $h_0$  depends on  $F$  through  $C_2$  and hence  $h_0$  must be estimated based on the random sample  $X_1, \dots, X_n$ .

Fortunately, the bandwidth given in Equation (8) has the same form as the optimal bandwidth that is used in the kernel estimation of continuous distribution functions. The only difference is the factor  $(1 - \alpha)$ . Note that  $n(1 - \alpha)$  is simply the expected number of observations in  $\mathbb{R} \setminus A$ , and so we can interpret the smoothed part of the estimate as having an effective sample of size  $n(1 - \alpha)$  instead of  $n$ . Further, if we estimate  $\alpha$  using Equation (1), then  $n(1 - \hat{\alpha})$  is the observed number of observations in  $\mathbb{R} \setminus A$ . This allows us to consider estimating  $h_0$  by applying already established bandwidth selection techniques for kernel distribution functions to the  $n(1 - \hat{\alpha})$  observations in  $\mathbb{R} \setminus A$ . Many methods have been proposed to estimate this bandwidth. A cross-validation procedure is discussed by Sarda (1993) though Altman and Leger (1995) have recently questioned the performance of this method and have presented their own results. A more reliable cross-validation procedure has recently been proposed by Bowman, Hall and Prvan (1998). Their method appears to provide substantially better performance than the method by Sarda (1993). Plug-in methods have been proposed by Altman and Leger (1995) and Polansky and Baker (2000). We will use the method of Polansky and Baker (2000) because a computer algorithm for the method is readily available in Polansky (2000). However, either of the methods proposed by Bowman, Hall and Prvan (1998) or Altman and Leger (1995) would also provide satisfactory performance. Indeed Bowman, Hall and Prvan (1998) conclude that any reasonable method, including a simple normal reference plug-in bandwidth estimator, will provide satisfactory results.

The method proposed by Polansky and Baker (2000) is a multistage plug-in bandwidth estimator. For the purpose of presenting the algorithm, let  $\tilde{n} = n(1 - \hat{\alpha})$  and  $\tilde{X}_1, \dots, \tilde{X}_{\tilde{n}}$  denote the observations in the sample that are in  $\mathbb{R} \setminus A$ . Suppose  $b > 0$  is an integer and let  $L$  be a kernel function, in this case a symmetric density with mean 0. Denote the  $j^{\text{th}}$  derivative of  $L$  as  $L^{(j)}$ . The function  $L$  need not be the same as  $k$ , but for simplicity we will use a standard normal density for  $L$  in the examples.

The algorithm for computing a  $b$ -stage plug-in estimator of  $h_0$  is outlined below.

Step 1. Calculate  $\widehat{\psi}_{2b+2}^{NR}$  where

$$\widehat{\psi}_r^{NR} = \frac{(-1)^{r/2} r!}{(2\widehat{\sigma})^{r+1} (r/2)! \pi^{1/2}},$$

where  $\widehat{\sigma}$  can be taken to be the sample standard deviation computed on  $\tilde{X}_1, \dots, \tilde{X}_{\tilde{n}}$ , or a measure suggested by Silverman (1986, pg. 47) that is more suitable for non-normal densities given by

$$\widehat{\sigma} = \min\{S, \text{IQR}/1.349\}, \quad (10)$$

where  $S$  is the sample standard deviation and IQR is the interquartile range computed on  $\tilde{X}_1, \dots, \tilde{X}_{\tilde{n}}$ .

Step 2. Starting with  $j = b$  and iterating until  $j = 1$ , calculate  $\widehat{\psi}_{2j}(\widehat{g}_{2j})$  where

$$\widehat{g}_{2j} = \left[ \frac{2L^{(2j)}(0)}{-\tilde{n}\mu_2(L)\widehat{\psi}_{2j+2}} \right]^{1/(2j+3)},$$

$$\widehat{\psi}_{2j+2} = \begin{cases} \widehat{\psi}_{2b+2}^{NR} & \text{when } j = b \\ \widehat{\psi}_{2j+2}(\widehat{g}_{2j+2}) & \text{when } j < b, \end{cases},$$

and

$$\widehat{\psi}_r(g) = \tilde{n}^{-2} g^{-r-1} \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} L^{(r)}\{(\tilde{X}_i - \tilde{X}_j)/g\}.$$

Step 3. Calculate

$$\widehat{h}_b = \left\{ \frac{C_1}{\tilde{n}\widehat{C}_2} \right\}^{1/3},$$

as the  $b$ -stage estimator of the optimal bandwidth where  $\widehat{C}_2 = -\mu_2^2(k)\widehat{\psi}_2(\widehat{g}_2)$ .

This type of rule has been used extensively in density estimation (see Chapter 3 of Wand and Jones (1995) for complete details) and more recently by Wand (1997) for selection of bin widths in histogram construction. The theoretical performance of this method, as outlined by Polansky and Baker (2000), follows its performance in histogram bin width selection given by Wand (1997). These studies suggest that a two-stage estimator ( $b = 2$ ) provides sufficient performance for most applications. The finite sample performance of the resulting estimator will closely follow that of the kernel smoothed distribution function estimator. See Polansky and Baker (2000) for an empirical study of the two-stage bandwidth estimator in that case.

### 3. EXAMPLES

#### 3.1 Tobacco and Alcohol Exposure in Animated Films

As an example consider the study by Goldstein, Sobel and Newman (1999) which presents two sets of data on alcohol and tobacco use in animated feature films. The first set of data reports the number of seconds showing alcohol use in each of 50 animated films released between 1937 and 1997. The second set of data reports the number of seconds of tobacco use in the same 50 animated films. As with many exposure data sets, there is a mass point in each of the data sets at 0 indicating films which did not show either alcohol or tobacco consumption. The remainder of the data can be taken to be continuous. A comparative plot of the empirical distribution function for each set of data is presented in Figure 1. The data reveals that 25 of the films did not show any alcohol consumption so that  $\hat{\alpha} = 0.50$  for the alcohol data set and that 22 of the films did not show any tobacco use so that  $\hat{\alpha} = 0.44$  for the tobacco data set. The two-stage bandwidth estimate calculated on the non-zero tobacco data was  $\hat{h}_0 = 40.3027$  and for the non-zero alcohol data was  $\hat{h}_0 = 26.3714$ . A comparative plot of the two smoothed estimates is given in Figure 2. One can observe from these plots that there are more films with tobacco use of longer length than of alcohol use. While the same general information is present in both plots, the trends in the smoothed plots are much easier to distinguish.

Note, however, that the kernel estimate has introduced some bias at the mass point 0. This is because the smoothed estimate does not account for the fact that the mass point 0 is also a boundary point for this problem, in the sense that  $F(x) = 0$  for  $x < 0$ . A similar problem occurs at boundaries in density estimation. In that case special kernel functions, call boundary kernels, can account for the boundary and significantly reduce the bias there. See, for example, Section 2.11 of Wand and Jones (1995) for examples of this type of behavior. It can be shown that similar boundary kernels can also be applied to the case of estimating distribution functions, though the resulting estimate will not always be non-decreasing. An alternative to the use of boundary kernels to solve this problem is to force the estimate of  $F(x)$  to be equal to  $\hat{\alpha}$  at 0, and then scale the remaining estimate so that  $\hat{F}(x; h) \rightarrow \infty$  as  $x \rightarrow \infty$ . This estimate has the form

$$\hat{F}_B(x; h) = \begin{cases} 0 & \text{if } x < 0, \\ \hat{\alpha} + (1 - \hat{\alpha})\{\hat{F}_2(x; h) - \hat{F}_2(0; h)\}/\{1 - \hat{F}_2(0; h)\} & \text{if } x \geq 0, \end{cases}$$

where  $\hat{F}_2(x; h)$  is the estimate given in Equation (3). This alternative has not been suggested for

use in density estimation because it does not reduce the order of the bias near the boundary. However, in the case of distribution functions it is well known that  $\hat{\alpha}$  is an unbiased estimator of  $\alpha$  so that the method is reasonable. This estimate is plotted for the two data sets in Figure 3. These plots are similar to those in Figure 2, except that the bias at the mass point 0 has now been removed.

### 3.2 Errors Arising in Audits

The second example investigates data consistent with that obtained from financial audits. We will specifically consider the errors detected by inventory and accounts receivable audits. In this type of data there is a discrete mass point at 0 indicating that there is a positive probability that a book value has no error. The remainder of the error distribution generally behaves in a continuous manner. Since actual audit data is difficult to obtain, we simulated 100 observations from distributions consistent with empirical error distributions observed in the study by Johnson, Leitch and Neter (1981). A comparative plot of the empirical distribution functions from these two simulated sets of data are presented in Figure 4. For the smoothed estimates we computed bandwidth estimates using the algorithm presented in Section 2. The corresponding smoothed estimates are plotted in Figure 5. Note again that it is easier to interpret the behavior in the smoothed plots. One can observe from Figure 5, for example, that aside from the different error rates, the continuous part of both distributions have the same general shape, but that the inventory audit error distribution has heavier tails. Moreover, both distribution function estimates indicate positive skewness.

## 4. DISCUSSION

A simple but effective smoothing method for nonparametrically estimating a nonstandard mixture distribution function was presented. This method is based on a mixture of an empirical distribution function and a smooth estimate of a distribution function. It was shown that the smoothed part of the estimator, constructed using a kernel estimate of a distribution function, has an optimal bandwidth that is the same as in the continuous case. Such a result implies that smoothing techniques that have already been developed for the continuous estimation of distribution functions can be used in the non-standard mixture case as well. A demonstration of the proposed method was exhibited using two examples of data that have a single mass point, but are otherwise continuous. In the case where the mass point occurs on a boundary, an estimate that reduces the bias incurred

at the mass point was introduced.

There are many opportunities for further work in this area. For example, confidence bands using the smoothed estimator can be developed using resampling techniques. To estimate a single point on the distribution function one can develop local smoothing methods similar to those used in density estimation or nonparametric regression. Finally, since the survival function is closely related to the distribution function, the development of smoothing methods in this case that account for censored observations could also be developed.

## APPENDIX

To obtain the result of Theorem 1, we first note that since  $F$  is a mixture of  $F_1$  and  $F_2$ , we can represent the random variable  $X \sim F$  as  $X = WD + (1 - W)C$  where  $D$ ,  $C$  and  $W$  are mutually independent random variables such that  $D \sim F_1$ ,  $C \sim F_2$ , and  $W$  is a Bernoulli random variable with success probability  $\alpha$ . Since  $X_1, \dots, X_n$  are independent and identically distributed random variables from  $F$ , we have that  $E[\hat{F}(x; h)] = E[\tilde{\Omega}_i(x; h)]$ . A well known property of conditional expectations yields

$$E[\tilde{\Omega}_i(x; h)] = (1 - \alpha)E[\tilde{\Omega}_i(x; h)|W = 0] + \alpha E[\tilde{\Omega}_i(x; h)|W = 1]. \quad (11)$$

The conditional distribution of  $F$  given  $W = 0$  is  $F_2$ , so that standard expansion theory yields (See Azzalini (1981) or Bowman, Hall and Prvan (1998)),

$$E[\tilde{\Omega}_i(x; h)|W = 0] = F_2(x) + h^2 c_2^{1/2}(x)/2 + o(h^2), \quad (12)$$

where  $c_2(x) = \mu_2^2(k)\{F_2''(x)\}^2$ . Similarly, the conditional distribution of  $F$  given  $W = 1$  is  $F_1$ , and since the empirical distribution function is pointwise unbiased,

$$E[\tilde{\Omega}_i(x; h)|W = 1] = F_1(x). \quad (13)$$

Note that we have also used the fact that the set  $A$  is known to derive Equations (11) - (13). Now, combining (11) - (13) yields

$$E[\hat{F}(x; h)] = F(x) + (1 - \alpha)h^2 c_2^{1/2}(x)/2 + o(h^2), \quad (14)$$

so that the square pointwise bias of  $\hat{F}(x; h)$  is

$$\text{Bias}^2[\hat{F}(x; h), F(x)] = h^4(1 - \alpha)^2 c_2(x)/4 + o(h^4). \quad (15)$$

To obtain the pointwise variance we note that

$$\text{Var}[\hat{F}(x; h)] = n^{-1} \{E[\tilde{\Omega}_i^2(x; h)] - E^2[\tilde{\Omega}_i(x; h)]\}, \quad (16)$$

where

$$E[\tilde{\Omega}_i^2(x; h)] = (1 - \alpha)E[\tilde{\Omega}_i^2(x; h)|W = 0] + \alpha E[\tilde{\Omega}_i^2(x; h)|W = 1]. \quad (17)$$

Using the relevant expansion theory from kernel distribution function estimates yields

$$E[\tilde{\Omega}_i^2(x; h)|W = 0] = F_2(x) - hF_2'(x)C_1 + o(h), \quad (18)$$

where

$$C_1 = \int_{-\infty}^{\infty} K(x)\{1 - K(x)\}dx.$$

Standard results from empirical distribution function theory yields

$$E[\tilde{\Omega}_i^2(x; h)|W = 1] = F_1(x). \quad (19)$$

Combining Equations (17) - (19) yields

$$E[\tilde{\Omega}_i^2(x; h)] = F(x) - h(1 - \alpha)F_2'(x)C_1 + o(h).$$

Now, Equation(14) implies that  $E^2[\tilde{\Omega}_i(x; h)] = F^2(x) + o(h)$ , so that the pointwise variance is given by

$$\text{Var}[\hat{F}(x; h)] = n^{-1}[F(x)\{1 - F(x)\} - h(1 - \alpha)F_2'(x)C_1] + o(n^{-1}h). \quad (20)$$

Finally, combining Equations (15) and (20) and integrating with respect to  $x$  yields

$$\text{MISE}[\hat{F}, F] = n^{-1} \int_{-\infty}^{\infty} F(x)\{1 - F(x)\}dx - n^{-1}h(1 - \alpha)C_1 + h^4(1 - \alpha)^2C_2/4 + o(h^4 + n^{-1}h).$$

## REFERENCES

- Aitchison, J. (1955) On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, **50**, 901-908.
- Altman, N. and Leger, C. (1995) Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, **46**, 195-214.
- Azzalini, A. (1981) A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326-328.
- Bowman, A., Hall, P. and Prvan, T. (1998) Bandwidth selection for the smoothing of distribution functions. *Biometrika*, **85**, 799-808.
- Chung, K. L. (1974) *A Course in Probability Theory*, Second Edition. Boston: Academic Press.
- Goldstein, A. O., Sobel, R. A. and Newman, G. R. (1999) Tobacco and alcohol use in G-rated children's animated films. *Journal of the American Medical Association*, **281**, No. 12, 1131-1136.
- Johnson, J. R., Leitch, R. A. and Neter, J. (1981) Characteristics of errors in accounts receivable and inventory audits. *The Accounting Review*, **56**, 270-293.
- Jones, M. C. (1990) The performance of kernel density functions in kernel distribution function estimation. *Statistics and Probability Letters*, **9**, 129-132.
- Nadaraya, E. A. (1964) Some new estimates for distribution functions. *Theory of Probability and its Applications*, **9**, 497-500.
- Panel on Nonstandard Mixtures of Distributions (1989) Statistical models and analysis in auditing. *Statistical Science*, **4**, 2-33.
- Polansky, A. M. (2000) An algorithm for computing a smooth nonparametric process capability estimate. *Journal of Quality Technology*, **32**, 284-289.
- Polansky, A. M. and Baker, E. R. (2000) Multistage plug-in bandwidth selection for kernel distribution function estimates. *Journal of Statistical Computation and Simulation*, **65**, 63-80.

- Rubin, B. K. (1990) Exposure of children with cystic fibrosis to environmental tobacco smoke. *The New England Journal of Medicine*, **323**, No. 12, 785.
- Sarda, P. (1993) Smoothing parameter selection for smooth distribution functions. *Journal of Statistical Planning and Inference*, **35**, 65-75.
- Schucany, W. R. and Wang, S. (1994) Bootstrap methods for nonstandard mixtures. In *Proceedings of the Section on Physical and Engineering Sciences*, pp. 59-64. Alexandria, VA: American Statistical Association.
- Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons.
- Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Vännman, K. (1995) On the distribution of the estimated mean from nonstandard mixtures of distributions. *Communications in Statistics - Theory and Methods*, **24**, 1569-1584.
- Wand, M. P. (1997) Data-based choice of histogram bin width. *The American Statistician*, **51**, 59-64.
- Wand, M. P. and Jones, M. C. (1995) *Kernel Smoothing*. London: Chapman and Hall.

Figure 1: Estimates of the distribution functions for the length of tobacco use (solid line) and alcohol use (dashed line) in the 50 animated films using the empirical distribution function.

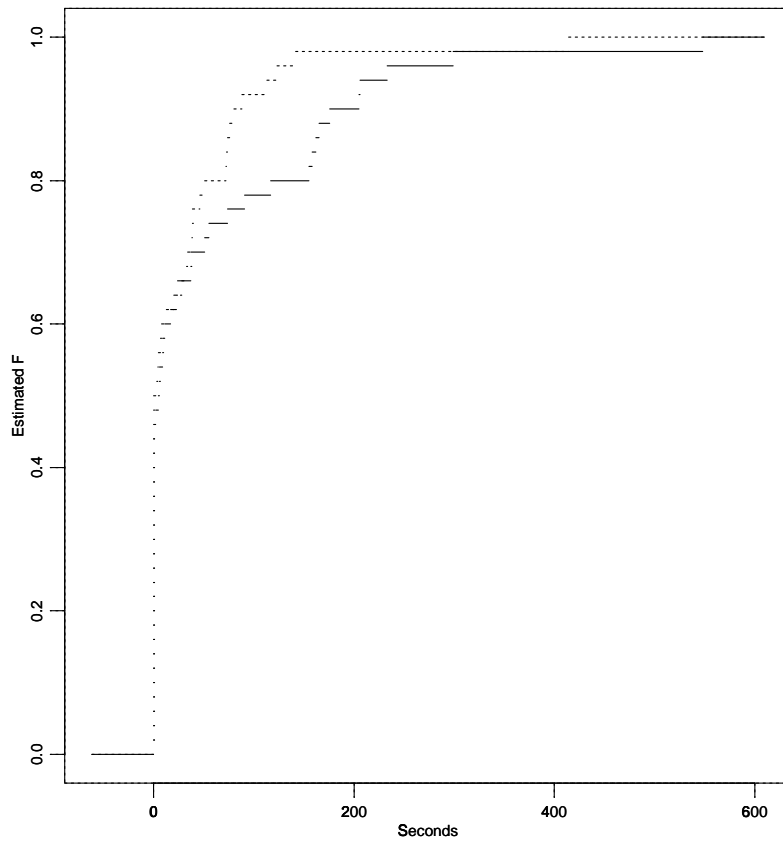


Figure 2: Estimates of the distribution functions for the length of tobacco use (solid line) and alcohol use (dashed line) in the 50 animated films using the smoothed estimate.

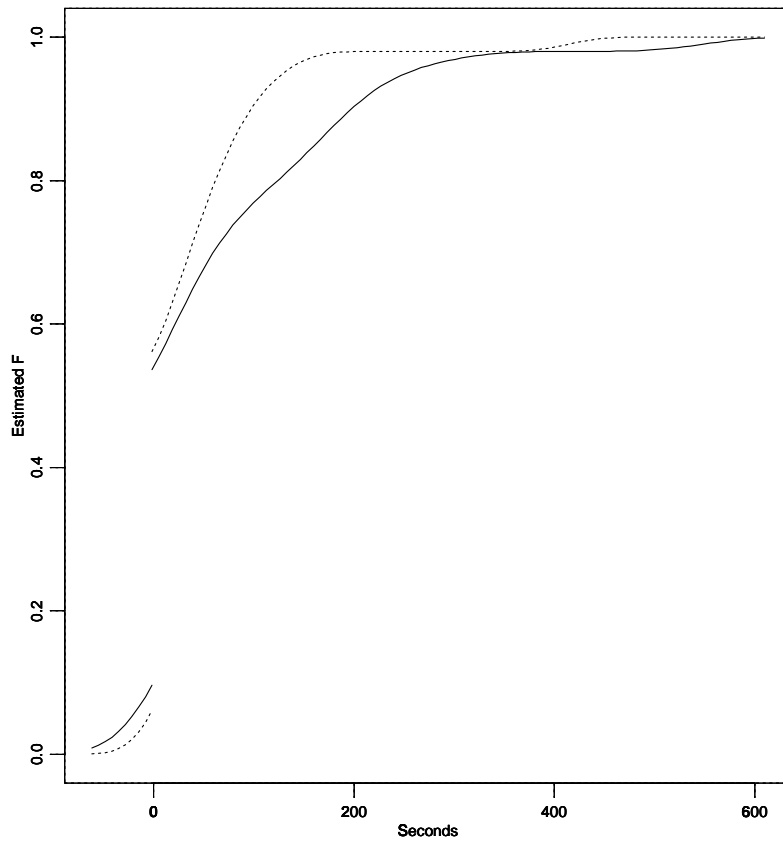


Figure 3: Estimates of the distribution functions for the length of tobacco use (solid line) and alcohol use (dashed line) in the 50 animated films using the smoothed estimate with boundary correction.

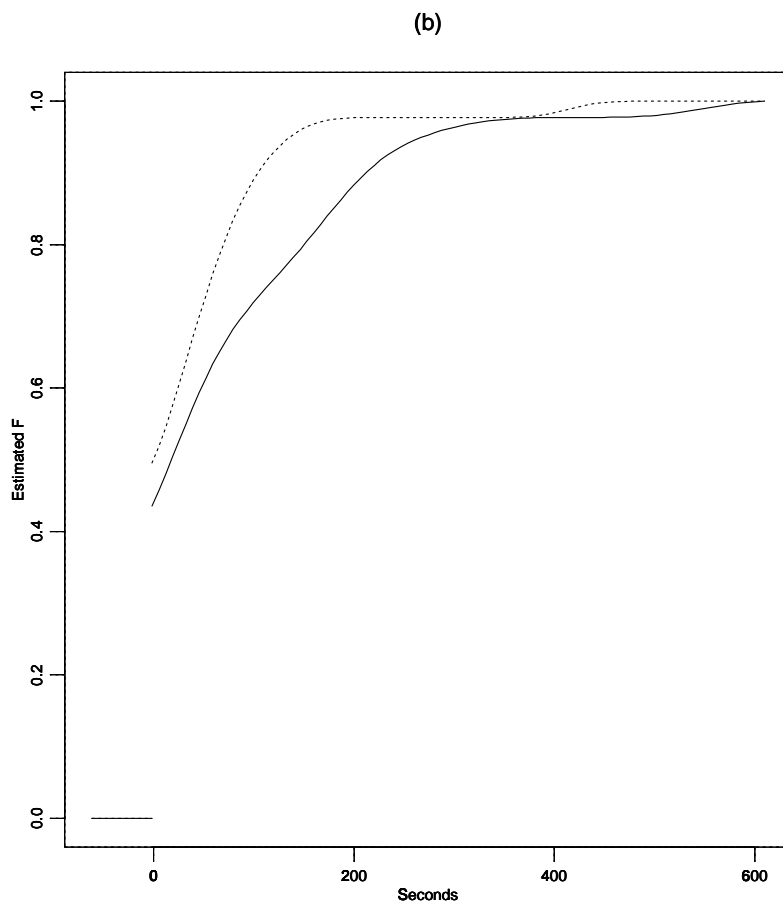


Figure 4: Estimates of the distribution functions for errors uncovered by an inventory audit (solid line) and an accounts receivable audit (dashed line) using the empirical distribution function.

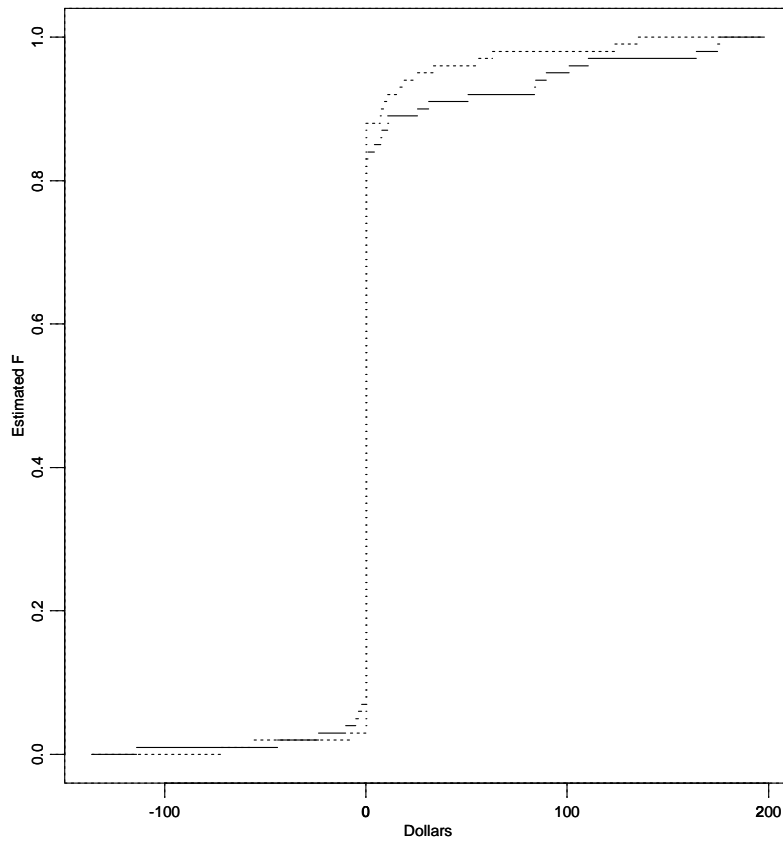


Figure 5: Estimates of the distribution functions for errors uncovered by an inventory audit (solid line) and an accounts receivable audit (dashed line) using the smoothed estimate.

